AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE

10.1002/2015WR017327

# A Bayesian hierarchical approach to model seasonal algal variability along an upstream to downstream river gradient

YoonKyung Cha[1], Seok Soon Park[2], Hye Won Lee[3], and Craig A. Stow[4]

[1]Department of Environmental Engineering, University of Seoul, Seoul, South Korea, [2]Department of Environmental Science and Engineering, College of Engineering, Ewha Womans University, Seoul, South Korea, [3]Center for Climate/Environment Change Prediction Research, Ewha Womans University, Seoul, South Korea, [4]NOAA Great Lakes Environmental Research Laboratory, Ann Arbor, Michigan, USA

**Key Points:**
- Chlorophyll is determined by river flow and phosphorus concentration
- The effects of flow and phosphorus concentration vary seasonally
- The effects of flow and phosphorus concentration differ upstream to downstream

**Supporting Information:**
- Supporting Information S1
- Supporting Information S2

**Correspondence to:**
C. A. Stow,
craig.stow@noaa.gov

**Citation:**
Cha, Y. K., S. Soon Park, H. Won Lee, and C. A. Stow (2016), A Bayesian hierarchical approach to model seasonal algal variability along an upstream to downstream river gradient, *Water Resour. Res.*, *52*, 348–357, doi:10.1002/2015WR017327.

Received 1 APR 2015
Accepted 21 DEC 2015
Accepted article online 28 DEC 2015
Published online 21 JAN 2016

**Abstract** Modeling to accurately predict river phytoplankton distribution and abundance is important in water quality and resource management. Nevertheless, the complex nature of eutrophication processes in highly connected river systems makes the task challenging. To model dynamics of river phytoplankton, represented by chlorophyll *a* (Chl *a*) concentration, we propose a Bayesian hierarchical model that explicitly accommodates seasonality and upstream-downstream spatial gradient in the structure. The utility of our model is demonstrated with an application to the Nakdong River (South Korea), which is a eutrophic, intensively regulated river, but functions as an irreplaceable water source for more than 13 million people. Chl *a* is modeled with two manageable factors, river flow, and total phosphorus (TP) concentration. Our model results highlight the importance of taking seasonal and spatial context into account when describing flow regimes and phosphorus delivery in rivers. A contrasting positive Chl *a*-flow relationship across stations versus negative Chl *a*-flow slopes that arose when Chl *a* was modeled on a station-month basis is an illustration of Simpson's paradox, which necessitates modeling Chl *a*-flow relationships decomposed into seasonal and spatial components. Similar Chl *a*-TP slopes among stations and months suggest that, with the flow effect removed, positive TP effects on Chl *a* are uniform regardless of the season and station in the river. Our model prediction successfully captured the shift in the spatial and monthly patterns of Chl *a*.

## 1. Introduction

Eutrophication, resulting principally from nonpoint nutrient inputs, is an ongoing problem in aquatic ecosystems worldwide [*Carpenter et al.*, 1998]. While the main cause of eutrophication is well understood, the problem has been difficult to control because nutrients originate from activities over broad areas and their inputs vary due to fluctuating weather and climatic conditions. Spatially and temporally varying conditions often induce considerable uncertainty in assessing management actions necessary to reduce eutrophication symptoms.

Numerical models that quantify the relationship between nutrients and appropriate eutrophication response indicators, as well as the uncertainty in this relationship, are essential tools to evaluate the outcome of alternative nutrient reduction strategies, and Bayesian networks offer a modeling framework that is particularly useful to support decision-making under uncertainty [*Kelly et al.*, 2013]. These models have been developed for various eutrophication response indicators in lakes and estuaries [*Borsuk et al.*, 2004; *Cha and Stow*, 2014; *Gudimov et al.*, 2012; *Nojavan et al.*, 2014; *Rigosi et al.*, 2015] but have not been widely applied in rivers.

Large river systems offer some particular opportunities for eutrophication model development. From upstream to downstream, rivers experience both gradual changes from nonpoint inputs reflective of the local landscape, as well as more sudden impacts from tributary and point-source inputs. Additionally, their advective nature results in a gradient of spatial and temporal integration from upstream to downstream, with upstream locations more reflective of shorter-term, local factors, and downstream locations expressing cumulative influences including the processing of inputs that occurs en route. Thus, conditions at locations along the upstream-downstream gradient may differ considerably.

From a modeling perspective, this broad range in conditions can be advantageous; generally a wider range in the available data will result in more precise model parameter estimation. Thus, one approach for parameter estimation would be to use all the data to estimate a set of model parameters that are assumed to be common across all sites. "Complete-pooling" was widely applied in early water quality modeling applications where data from many lakes or many sampling locations within a lake were used to estimate a single set of model parameters [*Canfield and Bachmann*, 1981; *Dillon and Rigler*, 1974; *Reckhow*, 1988].

While assuming a common parameter set may be appropriate for some applications, many factors can cause the relationship between model response and predictor variables to differ among sites. Thus another tactic would be to estimate a separate model for each site. "No pooling," however, fails to take advantage of the wide data range across sites and may produce site-specific parameter estimates with high uncertainty.

"Partial pooling" is a compromise between complete pooling and no pooling [*Gelman and Hill*, 2007]. In a Bayesian hierarchical modeling framework, partial pooling is accomplished by loosely constraining parameter sets across locations with a common prior distribution. Partial pooling can be particularly useful when parameter sets for some spatiotemporal combinations are poorly determined, either because they are based on a small sample or have a large variance. Poorly determined parameters are pulled toward the overall parameter mean, in a manner similar to that of "shrinkage estimators" [*Qian et al*., 2015; *Stow et al*., 2009].

The Nakdong River (Figure 1), located in the southeastern region of the Korean peninsula (35°N–37°N, 127°E–129°E), is an important water resource for agriculture, industry, and municipalities in southeastern Korea [*Park and Lee*, 2002]. Currently, about 7 million people reside within the basin and more than 13 million people obtain drinking water from the river. The Nakdong River is 525 km long and drains approximately 23,800 km$^2$. Of the total watershed area, 2.0% is residential and industrial, 17.0% is agricultural and paddy land, 69.0% is forest, and the remainder is classified as water and wetlands. The average annual temperature and total annual precipitation were 13.3°C and 1265 mm, respectively, during the years 2003–2012 (Korea Meteorological Administration, http://www.kma.go.kr). Precipitation during the summer (July to September) constituted 58% of the total annual precipitation.

During the past half century, a pronounced population increase associated with urban and industrial growth along the river has had negative impacts on the downstream water quality. In addition, because of the construction of an estuary barrier at the mouth of the river and four upstream dams, the river flow slowed and downstream areas exhibit characteristics similar to reservoirs. Since these alterations, eutrophication symptoms, including the excessive growth of bloom forming algal species, have become a critical issue [*Jeong et al*., 2006, 2007].

Despite these problems, the need for the Nakdong River's water resources is intense. In the mid-1990s, the construction and renovation of municipal wastewater facilities as well as stricter regulation on effluent discharges led to the improvement of several water quality indicators including biological oxygen demand (BOD) and dissolved oxygen (DO) [*Lee et al*., 2010]. However, phytoplankton blooms have continued to occur [*Kim et al*., 2007b]. More recently, eight weirs were constructed across the river as part of the Four Major Rivers Restoration Project.

To appropriately manage the river water quality and resources, development of a model that characterizes phytoplankton distribution and abundance, and identifies the factors that affect algal abundance is compelling. Previous models indicated that phytoplankton in the Nakdong River were influenced by dissolved nutrients [*Jeong et al*., 2001], temperature [*Kim et al*., 2007a,b], light availability [*Jeong et al*., 2001], evaporation [*Han et al*., 2009], or hydrology, including rainfall [*Jeong et al*., 2011; *Kim et al*., 2007b], dam discharge, and river flow [*Ha et al*., 2002; *Jeong et al*., 2006, 2007; *Kim et al*., 2007a; *Lee et al*., 2010]. In the Nakdong River pollutant levels, eutrophication symptoms and flows show an apparent upstream to downstream spatial gradient. Also, the Asian monsoon, characterized by hot, wet summer conditions, induces contrasting temperature and precipitation conditions to cold, dry winters. Thus far, no models have simultaneously accounted for both spatial and temporal variations in phytoplankton dynamics for the Nakdong River. To capture both the spatial and temporal variability in Nakdong River conditions, we developed a Bayesian hierarchical model (BHM) that predicts phytoplankton, measured as chlorophyll *a* concentration (Chl *a*). The approach enabled us to capture the spatial and temporal variability in Chl *a* levels, while also estimating the spatiotemporal differences in the relationships between Chl *a* and its predictors.

10.1002/2015WR017327

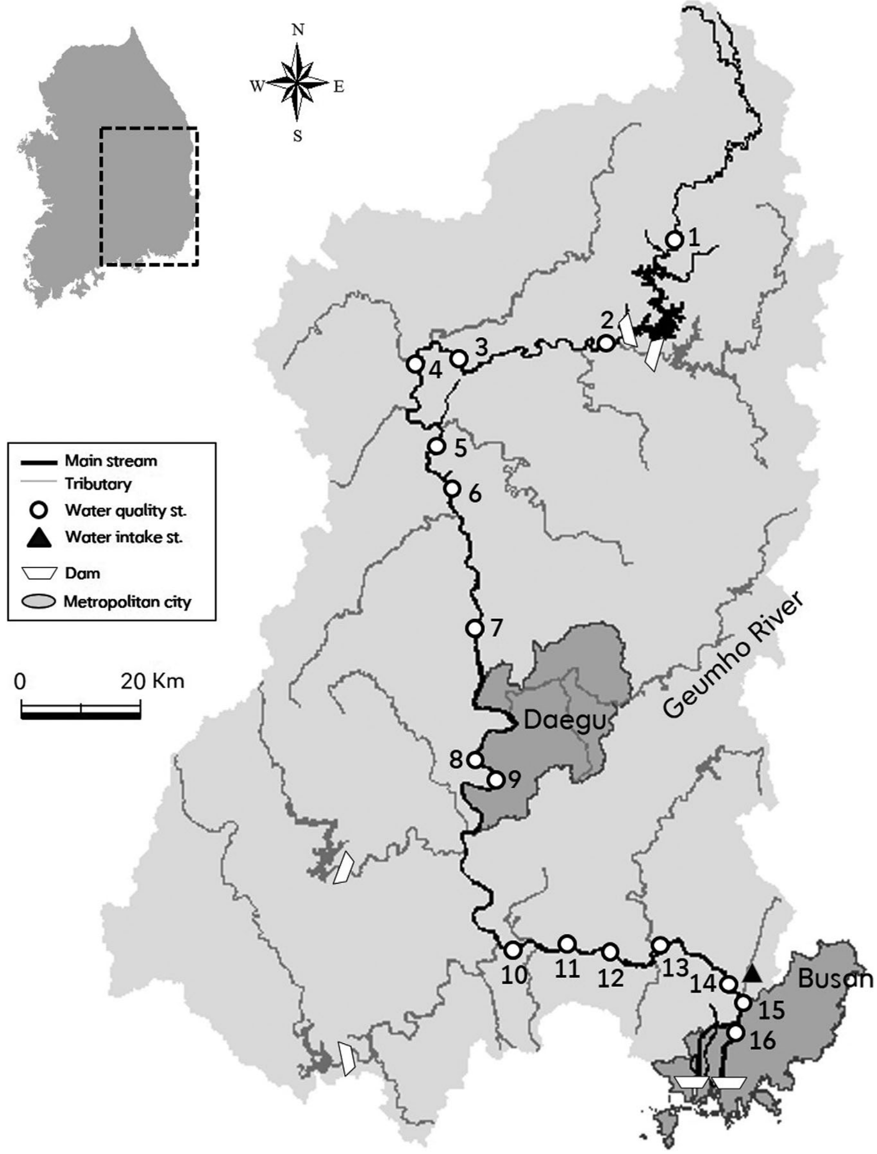


**Figure 1.** Map of the Nakdong River. Dams and estuarine barrages are marked with trapezoid. Sampling sites are denoted as 1–16 from the upper to lower parts of the river.

## 2. Methods

### 2.1. Data Description

We obtained data for phytoplankton and water quality characteristics at 16 monitoring stations monitored by National Institute of Environmental Research along the main stream (Figure 1). All samples were collected all year round (January to December) from 2003 to 2009. The study period was restricted to the years prior to 2010 when the Four Major Rivers Restoration Project began so that the predicted Chl *a* patterns are not confounded by the effects of the additional river engineering. Our intent is to provide a reference to evaluate the changes in phytoplankton dynamics corresponding to the construction of the weirs. The sampling frequency was $\sim$weekly at stations 7, 8, 10, and 16, and $\sim$monthly at the rest stations. Our data set included the following water quality characteristics which are expected to be related to phytoplankton dynamics: Chl *a*, total phosphorus (TP), and total nitrogen (TN). Daily precipitation and flow data were obtained from online WAMIS of the Korean Ministry of Land, Transport, and Maritime Affairs (http://www.wamis.go.kr). Daily irradiance was obtained from the Korea Meteorological Administration (http://www.kma.go.kr). The data used in this analysis are provided in the supporting information.

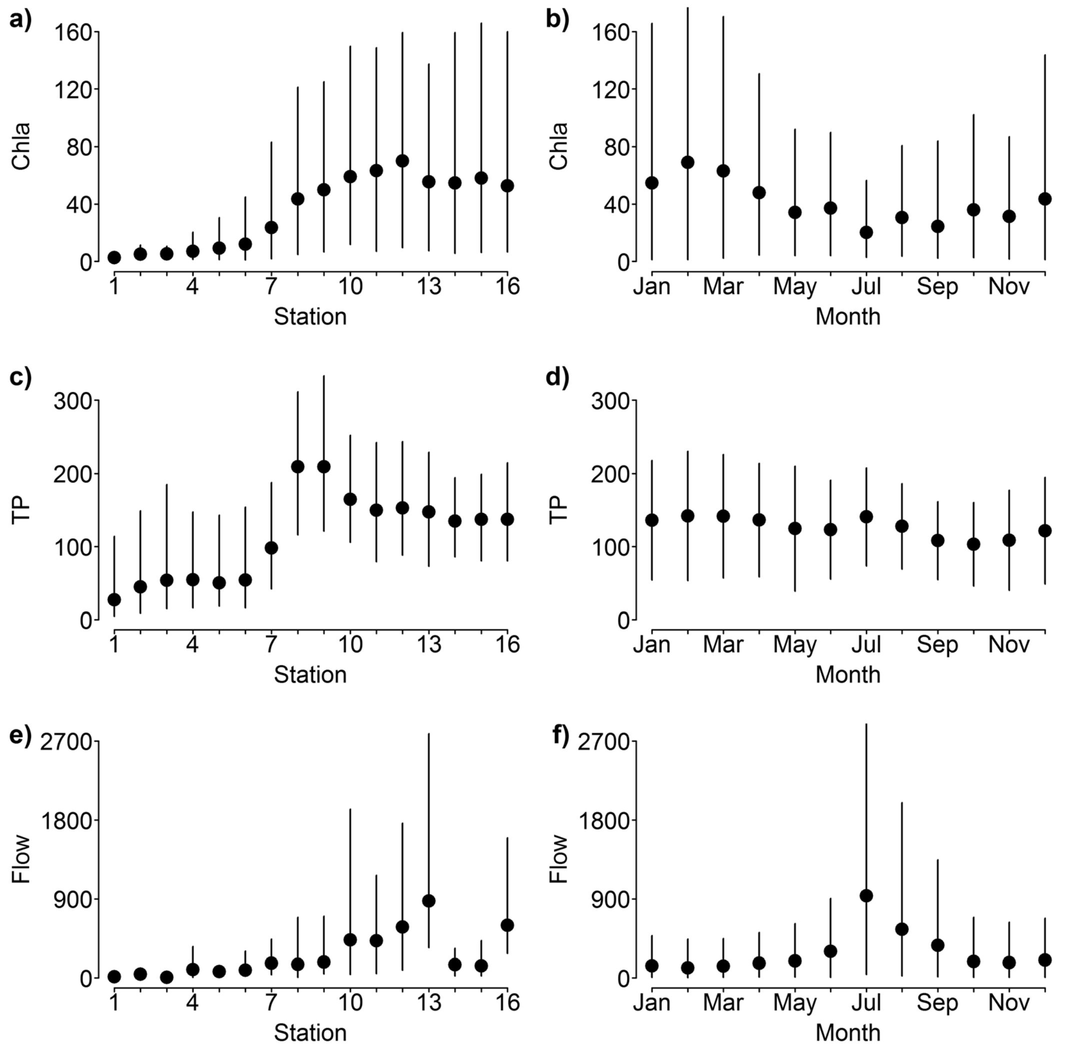


**Figure 2.** Spatial and monthly patterns of Chl *a* (μg/L), TP (μg/L), and rate of flow ($m^3/s$) in the Nakdong River. Circles and vertical bars indicate mean values, 5% and 95% percentiles.

### 2.2. Model Development

As a preliminary analysis, we examined relationships between Chl *a* and predictor candidates—TP, TN, flow, daily irradiance, daily precipitation, and 7 day total precipitation. Based on correlation and visual inspection, we selected TP and flow as predictors, which are manageable and consistently showed significant relationships with Chl *a* across months and stations. Further, we examined the pairwise relationships between log Chl *a* and log TP, log Chl *a* and log flow using both a simple ordinary least squares regression (complete pooling), and a BHM that allowed different intercepts and slopes for each station-month combination (partial pooling). We also examined the relationship between log TP and log flow to assess the potential effects of colinearity if both variables were included as predictors in the Chl *a* model.

Our final model was a BHM, that included both log TP and log flow as predictors, with intercepts and slopes that differed by group. We defined groups as a combination of station and month for a total of 192 (16 stations times 12 months). The response variable, Chl *a* concentration, was assumed to be lognormally distributed:

$$\log(\mathrm{Chl}a_i) \sim \mathrm{Normal}(\beta_{0j,k} + \beta_{1j,k} \cdot \log(TP_i) + \beta_{2j,k} \cdot \log(Flow_i), \sigma_y{}^2) \quad (1)$$

where $\mathrm{Chl}a_i$ is $i$th sample of Chl*a* concentration, which occurs in the $j$th station ($j = 1, \ldots, 16$), and $k$th month ($k$ = January, . . ., December). $\beta_0$ is intercept, $\beta_1$ and $\beta_2$ represent slope parameters. $\sigma_y{}^2$ is variance at

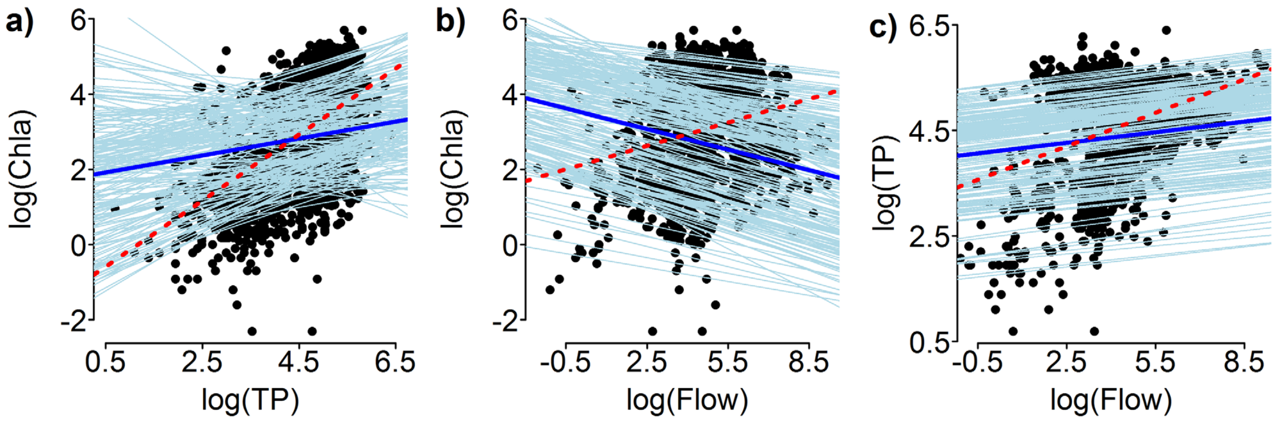


**Figure 3.** Exploratory analysis of the relationships between each pair of variables included in the model, Chl *a* (μg/L), TP (μg/L), flow ($m^3$/s), all on log scales. An ordinary least squares regression, based on complete-pooling of the data (red, dashed line) in each plot depicts the overall correlation across all observations (black circles). BHMs with distinct intercepts and slopes for each station-month combination are shown by the dark and light blue lines. The overall mean parameters of each BHM are shown by the dark blue lines, and the light-blue lines indicate each station-month combination.

the individual observation level. For each set of regression parameters, a probability distribution was assigned:

$$\beta_{m,j,k} \sim \text{Normal}\left(\mu_{\beta m}, \sigma_{\beta m}{}^2\right), \text{ for } m = 0, 1, \text{ and } 2$$

with their overall mean $\mu_\beta$ and standard deviation $\sigma_\beta$. We assumed noninformative prior distributions for all model parameters. A diffuse normal distribution was assigned for $\mu_\beta$ with a mean of zero and variance of $10^4$. Further we assigned uniform prior distributions with large intervals (1–100) for $\sigma_y$ and $\sigma_\beta$. Simulation of parameter posterior distributions were performed using Markov chain Monte Carlo (MCMC) procedures in the software program WinBUGS [*Lunn et al.*, 2000]. $10^5$ iterations of Gibbs sampling were produced with the first half discarded and the indication of model convergence was achieved with the potential-scale reduction parameter ($\hat{R}$) for all parameters equaled one.

## 3. Results

Overall average Chl *a* and TP concentrations were 40.5 μg (±0.9) and 125.7 μg (±1.5), respectively, over the years 2003–2009 (± denotes one standard error), indicating the eutrophic to hypereutrophic state of the river [*Carlson and Simpson*, 1996]. Chl *a* concentrations showed substantial differences with both space and time (Figures 2a and 2b). Chl *a* concentrations increased going downstream, remaining high at the lower part of the river (Figure 2a). Chl *a* concentrations were lower during the summer (July to September) than other seasons (Figure 2b). Similar to the Chl *a* patterns, TP concentrations exhibited a low to high spatial gradient from the upstream to downstream parts of the river. In contrast to Chl *a*, however, TP levels increased rapidly and peaked at stations 8 and 9 (Figure 2c), where the third largest city in South Korea, Daegu, is located (Figure 1). For TP concentration, seasonal patterns were not as distinct as those of Chl *a*, and exhibited a slight peak in July (Figure 2d), the time at which Chl *a* exhibited a minimum (Figure 2b).

Flow increased moving downstream, but decreased at stations 14 and 15, near to the river mouth (Figure 2e), where a large withdrawal provides water for the second largest city of South Korea, Busan (Figure 1). Seasonally, the flow was high during the summer monsoon season, peaking in July (Figure 2f).

Across all stations and months, log Chl*a* concentrations were positively correlated with log TP (Pearson correlation coefficient, $\rho = 0.58$, $n = 2385$) and flow ($\rho = 0.24$, $n = 1888$) (Figures 3a and 3b). Similarly, the correlation between log TP and log flow was positive ($\rho = 0.40$, $n = 2000$) (Figure 3c). However, further exploratory analysis of each pairwise relationship using a BHM with distinct intercepts and slopes for each station-month combination changed the picture considerably. Of the 192 estimated slopes for the log Chl *a* versus log TP model, 36 were negative and the overall slope of the BHM was 0.23, considerably less than that of the complete pooling, ordinary least squares regression line (Figure 3a). BHM results for the log Chl *a* versus log flow model revealed that the overall slope was negative as were 192 station-month slopes, a

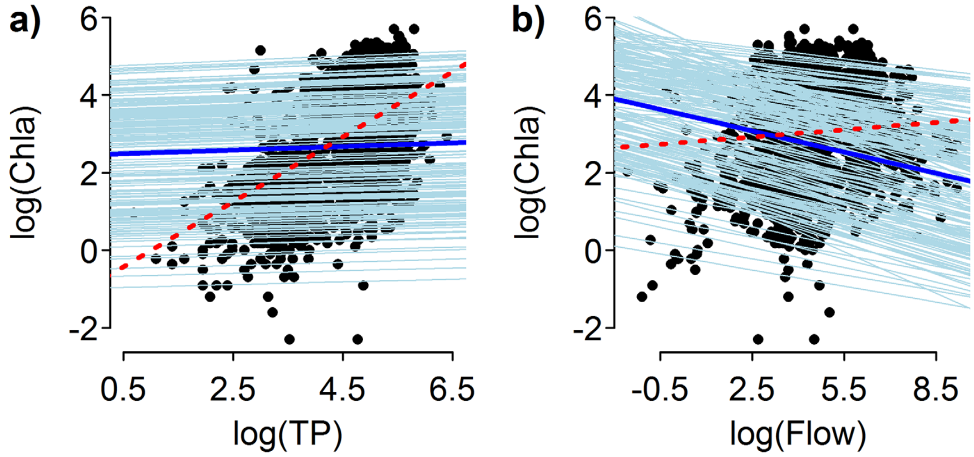


**Figure 4.** Model results of a response variable on the *y* axis with multiple predictors, TP and flow, on the *x* axis. Each plot displays between (a) Chl *a* (μg/L) and TP (μg/L) and (b) Chl *a* (μg/L) and flow ($m^3$/s) with all variables on logarithmic scales. Circles denote observations, dark blue lines indicate the Bayesian hierarchical regression line from overall mean parameters ($\mu_\alpha$ and $\mu_\beta$), and light blue lines indicate the hierarchical regression line for each station and month. Red dashed line indicates the complete-pooling regression line for all stations and months.

sharp contrast with positive slope resulting from complete pooling of the observations (Figure 3b). A BHM with log TP and log flow as the response and predictor variables, respectively, indicated consistent, and fairly similar positive slopes, of which all 192 were less than the complete-pooling slope (Figure 3c).

Results from the final model (Figure 4) which included both log TP and log flow as predictors of log Chl *a* (equation (1)), looked similar to the exploratory results examining log TP and log flow as separate predictors of Chl *a* (Figure 3). The BHM log TP slopes were all positive and very similar, ranging from 0.03 to 0.06; all were considerably less than the complete-pooling slope estimate of 0.84; 191 of the 192 log flow slopes were negative, ranging from −0.39 to 0.06, again in strong contrast with the positive complete-pooling slope estimate.

A comparison of mean monthly predictions and observed sample averages, by station (Figure 5), reveals that Chl *a* seasonality differed from upstream to downstream and that the across-site seasonal pattern (Figure 2b) largely reflects the pattern at the downstream stations 10–16. While the downstream stations 10–16 generally exhibited maxima from January to March and minima around July, several of the upstream stations exhibited almost the reverse pattern, with relative maxima during May to August, or displayed no clear seasonal cycle. Despite these differences among stations, the model mean monthly predictions captured the patterns at each station, generally falling very close to the observed sample averages. Of the 192 station-month mean predictions, only the station 5 August prediction differed notably from the observed sample average (Figure 5). This discrepancy probably reflects a relatively small sample ($n = 4$) that included one observation at an unusually high flow.

## 4. Discussion

Our study illustrated the spatial and seasonal dynamics of Chl *a* along the main channel of Nakdong River during the years 2003–2009, the period before the Four Major Rivers Restoration Project was implemented. Our results indicated that the Chl *a* levels in the river were strongly associated with flow and TP. These are consistent with previous findings in which flow regime and nutrient loading were identified as the important controlling factors of phytoplankton productivity in regulated rivers [*Chícharo et al.*, 2006; *Koch et al.*, 2004].

Across all river stations, Chl *a* concentration and flow had a positive relationship (Figures 3b and 4b, red dashed lines), which reflects upstream to downstream TP and flow (Figures 2c and 2e) increases. However, this positive Chl *a*-flow relationship across stations contrasted the negative Chl *a*-flow slopes that arose when Chl *a* was modeled as a function of flow and TP on a station-month basis (Figure 4b, blue lines). The

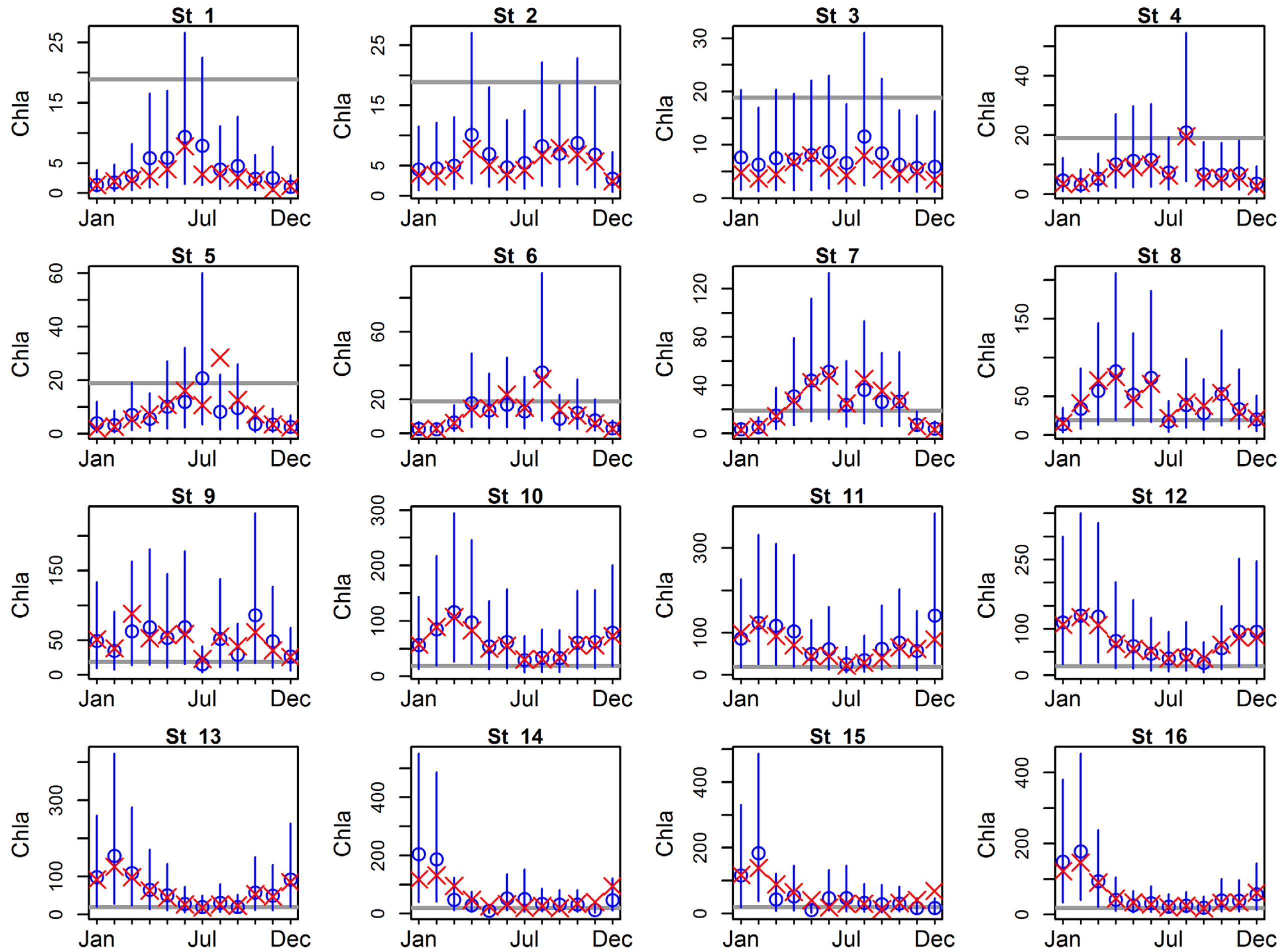


**Figure 5.** Predicted versus observed Chl *a* by station and month. Red crosses denote observed station-month mean values. Blue circles and vertical lines denote the predicted mean and 90% predictive interval at observed mean values. The gray horizontal line indicates the overall predicted mean across stations and months.

fact that this contrast occurs whether Chl *a* is modeled as a function of both TP and flow (Figure 4b) or only flow (Figure 3b) indicates that the sign reversal of the slopes does not result from collinearity between flow and TP (Figure 3c). Rather, this change is an illustration of Simpson's paradox [*Simpson*, 1951], which occurs when trends among observations differ depending on how the data are aggregated. Simpson's paradox can result in faulty inference if data are improperly grouped [*Bickel et al.*, 1975]. In this case, causal inference could be incorrect, with serious management consequences. Our results which captured the overall mean negative slope ($\mu_{\beta 2} = -0.18$) between Chl *a* and flow derived from within-group slopes ($\beta_2$) ranging from negative ($-0.39$) to slightly positive (0.06) suggest that hierarchical regression with partial pooling is an approach to avoid the pitfall of the Simpson's paradox.

Increases in river flow, precipitation, or dam outflow promote decreased residence time [*Chícharo et al.*, 2006; *Koch et al.*, 2004]. Flow effects on flushing, diluting, or preventing phytoplankton blooms in the Nakdong River were reported in previous studies. Concentrated precipitation in summer suppressed cyanobacteria blooms [*Ha et al.*, 2002; *Kim et al.*, 2007b] had a prolonged effect on reducing Chl *a* concentrations in the following season [*Jeong et al.*, 2011]. Also, *Ha et al.* [2003] and *Jeong et al.* [2011] suggested that the formation and magnitude of winter blooms in the Nakdong River were influenced by annual precipitation or dam discharge in winter. In other flowing systems, such as the Guadiana River, Portugal [*Chícharo et al.*, 2006], Murrumbidgee River, Australia [*Viney et al.*, 2007; *Webster et al.*, 2000], Hunter River, Australia [*Mitrovic et al.*, 2008], and Barwon-Darling River, Australia [*Mitrovic et al.*, 2011], the role of hydrological pulses in suppressing or promoting phytoplankton blooms was also reported to be critical. In these rivers, flow regulation by dam operation was an adequate and often sole management strategy for controlling the development and duration of nuisance blooms. *Mitrovic et al*. [2011] reported that a critical discharge of 300 million liter per day was required to prevent a cyanobacterial bloom from developing, and 3000 million liter per day was required to eliminate a developed bloom in the lower Darling River.

Upstream to downstream TP levels revealed systematic, noticeable changes in the Nakdong River. In stations 1–7, TP concentrations indicated a eutrophic state (mean = 58.4 μg/L), went up to the mean level of 209.1 μg/L in stations 8 and 9, and remained hypereutrophic in stations 10–16 (mean = 145.6 μg/L). The abrupt TP increase in stations 8 and 9 is attributable to the confluence of the Geumho River, where wastewater and effluent from Daegu City are discharged (Figure 1), and highlights the importance of input from this source in determining the nutrient level of the lower Nakdong River [*Han et al*., 2009]. *Yoon et al*. [2014] also pointed out that along the river, Daegu City is the single most important contributor to downstream eutrophication and deteriorated water quality.

Positive TP-flow relationships at all stations and months (Figure 3c) suggest that agricultural and urban runoff, over dilution, is a dominant mechanism controlling TP levels across the system [*Johnson*, 1979]. Aggressive point-source control in the mid-1990s [*Lee et al*., 2010] likely reduced the relative importance of point sources, and controlling nonpoint source of phosphorus should be a management priority to reduce eutrophication in the river.

Variable Chl *a*-TP slopes (Figure 3a) became positive, invariant across stations and months when Chl *a* was modeled with both flow and TP (Figure 4a). Also, note that between-group slope standard deviation for Chl *a*-TP ($\sigma_{\beta 1} = 0.04$), was close to zero, while the standard deviation for Chl *a*-flow ($\sigma_{\beta 2} = 0.14$) was much larger, meaning that grouping by station-month delivers little information in estimating Chl *a*-TP slopes. The changes in Chl *a*-TP relationships along with small group-level slope standard deviation for Chl *a*-TP suggest that, with the flow effect accounted for, TP effects on Chl *a* are similar among differing stations and months.

Because our hierarchical approach had a structure suitable to represent the spatial and temporal gradients in hydrological and nutritional conditions, the shifts in observed peaks of Chl *a* from site to site, and also from month to month were successfully captured in the model. These shifts suggest that the dominant type of phytoplankton and algal productivity have changed along the river over the course of a year. From station 1 to station 7, Chl *a* levels gradually increased and the summer peak during July to August became apparent. This pattern indicates as nutrient level increases the summer blooms dominates the annual phytoplankton production in midriver stations. Cyanobacteria, including *Microcystis*, *Aphanizomenon*, *Anabaena*, are known as dominant genera in the river during summer [*Ha et al*., 2002; *Hur et al*., 2013]. In stations 8 and 9, the summer Chl *a* peak collapsed and the transition to the winter peak (January to February) occurred in stations 10–16. Despite the hypereutrophic state, high summer flow might have suppressed cyanobacteria proliferation in these stations. The summer flow in stations 10–16 (mean = 1191.5 $m^3\ s^{-1}$ during July to August) was substantially greater than the summer flow in stations 1–7 (mean = 207.5 $m^3\ s^{-1}$). Another explanation for the summer peak disappearance in downstream stations is that the summer algal blooms in upstream stations led to phytoplankton-induced nutrient depletion in downstream stations of the river [*Koch et al*., 2004]. Previous studies reported that the winter dominance of diatoms, such as *Stephanodiscus hanzschii*, in the lower river requires low rainfall and river flow [*Ha et al*., 2003; *Jeong et al*., 2007]. Our data also indicated that the river flow during January to February in stations 10–16 (mean = 228.1 $m^3\ s^{-1}$) was substantially lower than the summer flow in those stations.

After the completion of the Four Major Rivers Restoration Project, which aimed at preventing water deficiency and flooding, eight weirs were constructed along the Nakdong River. The river is expected to exhibit more reservoir-like ecology and hydrology with increased residence time. The changes in seasonal and spatial characteristics of flow and residence time would substantially change the Chl *a*-flow relationships during the preweir construction period, also altering the suppressing effects of flow increases on summer and winter phytoplankton blooms. Continued monitoring efforts and future studies to reveal the spatiotemporal changes in Chl *a*-flow relationships and phytoplankton dynamics in the river after the construction of weirs are invited.

*Koch et al*. [2004] suggested that as river water changes from free-flowing toward standing due to dam construction, phytoplankton productivity tends to be nutrient limited and less light limited (light saturated). In our data set, Chl *a* showed a weak correlation with irradiance ($\rho = 0.098$), while showing generally strong correlations with nutrient indicators. An increase in nutrient sedimentation and retention subsequent to the recent construction of weirs is expected to loosen the link between nutrient input and algal blooms for the time being, but future eutrophication issues originating in sediment nutrient release and nutrient resuspension should also be considered.

## 5. Conclusion

Partial pooling in a Bayesian hierarchical modeling framework effectively predicted Chl *a* seasonal dynamics in the Nakdong River system, and revealed causal dependencies that will be useful in future management decision-making. Our analysis offered an interesting example of Simpson's paradox, which can occur when data are aggregated across variables that may be causally influencing the response variable [*Pearl*, 2000]. In this instance, both predictor variables, TP and flow, were confounded by factors associated with space (station) and time (month); thus, aggregating spatially and temporally suggested misleading relationships between Chl *a* and the predictor variables. The Bayesian hierarchical modeling framework controlled for these confounding effects, resulting in an informative, predictive model.

**Acknowledgments**
The data used in this analysis are provided in the supporting information. GLERL contribution 1792.

## References

Bickel, P. J., E. A. Hammel, and J. W. O'Connell (1975), Sex bias in graduate admissions: Data from Berkeley, *Science*, *187*(4175), 398–404.
Borsuk, M. E., C. A. Stow, and K. H. Reckhow (2004), A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis, *Ecol. Modell.*, *173*(2–3), 219–239.
Canfield, D. E., Jr., and R. W. Bachmann (1981), Prediction of total phosphorus concentrations, chlorophyll *a*, and Secchi depths in natural and artificial lakes, *Can. J. Fish. Aquat. Sci.*, *38*, 414–423.
Carlson, R. E., and J. Simpson (1996), *A Coordinator's Guide to Volunteer Lake Monitoring Methods*, North Am. Lake Manage. Soc., Madison, Wis.
Carpenter, S. R., N. F. Caraco, D. L. Correll, R. W. Howarth, A. N. Sharpley, and V. H. Smith (1998), Nonpoint pollution of surface waters with phosphorus and nitrogen, *Ecol. Appl.*, *8*(3), 559–568.
Cha, Y., and C. A. Stow (2014), A Bayesian network incorporating observation error to predict phosphorus and chlorophyll a in Saginaw Bay, *Environ. Modell. Software*, *57*, 90–100.
Chícharo, L., M. A. Chícharo, and R. Ben-Hamadou (2006), Use of a hydrotechnical infrastructure (Alqueva Dam) to regulate planktonic assemblages in the Guadiana estuary: Basis for sustainable water and ecosystem services management, *Estuarine Coastal Shelf Sci.*, *70*(1), 3–18.
Dillon, P. J., and F. H. Rigler (1974), The phosphorus-chlorophyll relationship in lakes, *Limnol. Oceanogr.*, *19*, 767–773.
Gelman, A., and J. Hill (2007), *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge Univ. Press, N. Y.
Gudimov, A., E. O'Connor, M. Dittrich, H. Jarjanazi, M. E. Palmer, E. Stainsby, J. G. Winter, J. D. Young, and G. B. Arhonditsis (2012), Continuous Bayesian network for studying the causal links between phosphorus loading and plankton patterns in Lake Simcoe, Ontario, Canada, *Environ. Sci. Technol.*, *46*, 7283–7292.
Ha, K., M.-H. Jang, and G.-J. Joo (2002), Spatial and temporal dynamics of phytoplankton communities along a regulated river system, the Nakdong River, Korea, *Hydrobiologia*, *470*(1–3), 235–245.
Ha, K., M.-H. Jang, and G.-J. Joo (2003), Winter *Stephanodiscus* bloom development in the Nakdong River regulated by an estuary dam and tributaries, *Hydrobiologia*, *506*(1–3), 221–227.
Han, S., E. Kim, and S. Kim (2009), The water quality management in the Nakdong River watershed using multivariate statistical techniques, *KSCE J. Civ. Eng.*, *13*(2), 97–105.
Hur, M., I. Lee, B.-M. Tak, H. J. Lee, J. J. Yu, S. U. Cheon, and B.-S. Kim (2013), Temporal shifts in cyanobacterial communities at different sites on the Nakdong River in Korea, *Water Res.*, *47*(19), 6973–6982.
Jeong, K.-S., G.-J. Joo, H.-W. Kim, K. Ha, and F. Recknagel (2001), Prediction and elucidation of phytoplankton dynamics in the Nakdong River (Korea) by means of a recurrent artificial neural network, *Ecol. Modell.*, *146*(1), 115–129.
Jeong, K.-S., D.-K. Kim, and G.-J. Joo (2006), River phytoplankton prediction model by Artificial Neural Network: Model performance and selection of input variables to predict time-series phytoplankton proliferations in a regulated river system, *Ecol. Inf.*, *1*(3), 235–245.
Jeong, K.-S., D.-K. Kim, and G.-J. Joo (2007), Delayed influence of dam storage and discharge on the determination of seasonal proliferations of *Microcystis aeruginosa* and *Stephanodiscus hantzschii* in a regulated river system of the lower Nakdong River (South Korea), *Water Res.*, *41*(6), 1269–1279.
Jeong, K.-S., D.-K. Kim, H.-S. Shin, J.-D. Yoon, H.-W. Kim, and G.-J. Joo (2011), Impact of summer rainfall on the seasonal water quality variation (chlorophyll a) in the regulated Nakdong River, *KSCE J. Civ. Eng.*, *15*(6), 983–994.
Johnson, A. H. (1979), Estimating solute transport in streams from grab samples, *Water Resour. Res.*, *15*(5), 1224–1228.
Kelly, R. A., A. J. Jakeman, O. Barreteau, M. E. Borsuk, S. ElSawah, S. H. Hamilton, H. J. Henriksen, S. Kuikka, H. R. Maier, and A. E. Rizzoli (2013), Selecting among five common modelling approaches for integrated environmental assessment and management, *Environ. Modell. Software*, *47*, 159–181.
Kim, D.-K., K.-S. Jeong, P. A. Whigham, and G.-J. Joo (2007a), Winter diatom blooms in a regulated river in South Korea: Explanations based on evolutionary computation, *Freshwater Biol.*, *52*(10), 2021–2041.
Kim, D.-K., H. Cao, K.-S. Jeong, F. Recknagel, and G.-J. Joo (2007b), Predictive function and rules for population dynamics of *Microcystis aeruginosa* in the regulated Nakdong River (South Korea), discovered by evolutionary algorithms, *Ecol. Modell.*, *203*(1), 147–156.
Koch, R. W., D. L. Guelda, and P. A. Bukaveckas (2004), Phytoplankton growth in the Ohio, Cumberland and Tennessee Rivers, USA: Inter-site differences in light and nutrient limitation, *Aquat. Ecol.*, *38*(1), 17–26.
Lee, H. W., K. J. Bhang, and S. S. Park (2010), Effective visualization for the spatiotemporal trend analysis of the water quality in the Nakdong River of Korea, *Ecol. Inf.*, *5*(4), 281–292.

Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter (2000), WinBUS—A Bayesian modelling framework: Concepts, structure, and extensibility G, *Stat. Comput.*, *10*, 325–337.
Mitrovic, S. M., B. C. Chessman, A. Davie, E. L. Avery, and N. Ryan (2008), Development of blooms of *Cyclotella meneghiniana* and *Nitzschia* spp. (Bacillariophyceae) in a shallow river and estimation of effective suppression flows, *Hydrobiologia*, *596*(1), 173–185.
Mitrovic, S. M., L. Hardwick, and F. Dorani (2011), Use of flow management to mitigate cyanobacterial blooms in the Lower Darling River, Australia, *J. Plankton Res.*, *33*(2), 229–241.
Nojavan, F., S. S. Qian, H. W. Paerl, K. H. Reckhow, and E. A. Albright (2014), A study of anthropogenic and climatic disturbance of the New River Estuary using a Bayesian belief network, *Mar. Pollut. Bull.*, *83*(1), 107–115.
Park, S. S., and Y. S. Lee (2002), A water quality modeling study of the Nakdong River, Korea, *Ecol. Modell.*, *152*(1), 65–75.
Pearl, J. (2000), *Causality*, Cambridge University Press, Cambridge, U. K.
Qian, S. S., C. A. Stow, and Y. Cha (2015), Implications of Stein's Paradox for environmental standard compliance assessment, *Environ. Sci. Technol.*, *49*(10), 5913–5920.
Reckhow, K. H. (1988), Empirical models for trophic state in southeastern United States lakes and reservoirs, *Water Resour. Bull.*, *24*, 723–734.
Rigosi, A., P. Hanson, D. P. Hamilton, M. Hipsey, J. A. Rusak, J. Bois, K. Sparber, I. Chorus, A. J. Watkinson, and B. Qin (2015), Determining the probability of cyanobacterial blooms: The application of Bayesian networks in multiple lake systems, *Ecol. Appl.*, *25*(1), 186–199.
Simpson, E. H. (1951), The interpretation of interaction in contingency tables, *J. R. Stat. Soc., Ser. B*, *13*(2), 238–241.
Stow, C. A., E. C. Lamon III, S. S. Qian, P. A. Soranno, and K. H. Reckhow (2009), Bayesian hierarchical/multilevel models for inference and prediction using cross-system lake data, in *Real World Ecology: Large-Scale and Long-Term Case Studies and Methods*, edited by S. Miao, S. Carstenn, and M. Nungesser, pp. 111–136, Springer, N. Y.
Viney, N. R., B. C. Bates, S. P. Charles, I. T. Webster, and M. Bormans (2007), Modelling adaptive management strategies for coping with the impacts of climate variability and change on riverine algal blooms, *Global Change Biol.*, *13*(11), 2453–2465.
Webster, I. T., B. S. Sherman, M. Bormans, and G. Jones (2000), Management strategies for cyanobacterial blooms in an impounded lowland river, *Regul. Rivers Res. Manage.*, *16*(5), 513–525.
Yoon, T., C. Rhodes, and F. A. Shah (2014), Upstream water resource management to address downstream pollution concerns: A policy framework with application to the Nakdong River basin in South Korea, *Water Resour. Res.*, *51*, 787–805, doi:10.1002/2013WR014201.